



**Diseño Factorial De Efectos Fijos 4X2 Para Evaluar La
Medición De Alcalinidad Total. Estudio De Caso**

Blanca Ceidy Clavijo Ortiz
Licenciada En Química

Madelyn Piñeros Gutiérrez
Licenciada En educación Básica énfasis en Matemáticas

Fundación Universitaria Los Libertadores
Departamento de Ciencias Básicas
Especialización en estadística aplicada

Bogotá D.C.
2016



**Diseño Factorial De Efectos Fijos 4X2 Para Evaluar La
Medición De Alcalinidad Total. Estudio De Caso**

Blanca Ceidy Clavijo Ortiz
Licenciada En Química

Madelyn Piñeros Gutiérrez
Licenciada En educación Básica énfasis en Matemáticas

Asesor estadístico: Martha Tatiana Pamela Jimenez Valderrama
Asesor metodológico: Diana Patricia Walteros

Fundación Universitaria Los Libertadores
Departamento de Ciencias Básicas
Especialización en estadística aplicada

Bogotá D.C.
2016

Nota de Aceptación

Firma del presidente del jurado

Firma del Jurado

Firma del Jurado

Bogotá, D.C 31 Julio del 2016

Las Directivas de la Universidad de
Los Libertadores, los jurados calificadores y el cuerpo
Docente no son responsables por los
criterios e ideas expuestas En el presente documento.
Estos corresponde únicamente a los autores

Contenido

Resumen	10
Capítulo 1. Introducción	11
Objetivo General.....	12
Objetivos Específicos	12
Capítulo 3. Marco Teórico	15
3.1 Diseño de experimentos	15
3.2 Función de un diseño experimental	15
3.2.1. Modelo unifactorial o de enfoque de un factor a la vez	16
3.2.2 Modelo factorial	18
3.3 Verificación de la adecuación del modelo	21
3.3.1 Supuesto de Normalidad.....	21
3.3.2 Homogeneidad de Varianzas (Homocedasticidad).....	21
3.4 Comparaciones múltiples.....	22
Capítulo 4. Marco Metodológico	23
4.1 Unidades experimentales:	23
4.2 Factores y niveles:.....	23
4.3 Variable Respuesta:	23
4.4 Tratamientos:.....	24
4.5 Procedimiento.....	24
Capítulo 5. Análisis y Resultados	26
5.1 Diagrama exploratorio (Gráfico de medias).....	26
5.2 ANOVA de un diseño con estructura factorial	26
5.3 Validación de supuestos	27
5.3.1 Normalidad.....	27
5.3.2 Homocedasticidad.....	28
5.4 Determinación del modelo	28
5.5 Prueba de comparaciones múltiples	29
5.6 Análisis Unifactorial	31
Capítulo 6. Conclusiones	34
Capítulo 7. Referencias	35

Anexos 37

Anexo 1 37

Anexo 2..... 37

Rutas empleadas en “R Commander” 37

Lista de Tablas

Tabla 1. Datos típicos de un experimento de un solo factor	16
Tabla 2. Tabla de análisis de varianza para el modelo con un solo factor y efectos fijos.	17
Tabla 3. Arreglo general de un diseño factorial de dos factores	19
Tabla 4. Tabla de Análisis de varianza multifactorial	20
Tabla 5. Valores de Alcalinidad Total en mg/L de CaCO_3	24
Tabla 6. Datos previos formación Diseño Factorial 4x2	24
Tabla 7. Resultados prueba Homocedasticidad de Factores Individuales	32

Lista de figuras

Figura 1. Gráfico de medias "Operador - Punto de Muestreo"	26
Figura 2. Tabla de ANOVA estructura factorial.....	27
Figura 3. Resultado Prueba Normalidad Shapiro - Wilk, Software R	27
Figura 4. Resultado Prueba Homocedasticidad Bartlett, Levene.	28
Figura 5. Prueba Tukey - Comparaciones múltiples	29
Figura 6. Intervalos de confianza de los tratamientos.....	30
Figura 7. Diagramas de caja. a) Punto de Muestreo. b) Operador	31
Figura 8. ANOVA de un solo factor para operador y punto de muestreo	32
Figura 9. Prueba Tukey para el factor operador. a) Contrastes. b) Intervalos de confianza.....	33

Glosario

Análisis de varianza: Refiere al análisis que prueba la hipótesis de que las medias de las poblaciones en comparación son iguales, teniendo en cuenta la evaluación de los factores con respecto a los distintos niveles.

Contraste: Se denomina contraste a toda combinación lineal de los parámetros del modelo de análisis de varianza.

Efectos fijos: Es un factor en el que los niveles han sido seleccionados por el experimentador. Es apropiado cuando el interés se centra en comparar el efecto sobre la respuesta de esos niveles específicos.

Error experimental: Refiere a la variación entre las unidades experimentales en concordancia con el tratamiento que se les haya dado.

Factores fijos: Son todas las fuentes de variación reconocidas a priori por el investigador (no incluye el error experimental). Cada uno de los valores que asume un factor se denomina nivel.

Niveles: Tipos o grados del factor tenidos en cuenta a la hora de realizar el experimento.

Replicas: Corridas experimentales con las mismas configuraciones de los niveles.

Tratamiento: Hace referencia a la selección de condiciones elegidas por el experimentador a la hora de realizar cada prueba.

Unidad experimental: Material donde se evalúa la variable respuesta y al que se le aplican los distintos niveles de los factores tratamiento.

Variable efecto: Es el factor que es manipulado por el experimentador durante el proceso, con la intención de saber cuál es el comportamiento de la variable respuesta de acuerdo con los diferentes niveles que se dispongan.

Variable Respuesta: Hace referencia a la variable de estudio, es decir, de la que se pretende observar la variabilidad al ser sometida a cada experimento.

DISEÑO FACTORIAL DE EFECTOS FIJOS 4X2 PARA EVALUAR LA MEDICIÓN DE ALCALINIDAD TOTAL. ESTUDIO DE CASO

****Blanca Ceidy Clavijo Ortiz, Madelyn Piñeros Gutiérrez**

Resumen

El análisis multifactorial se emplea a menudo para la evaluación de varios factores incidentes sobre una misma variable respuesta, para este caso particular se estudia la posible existencia de efectos cuando hay variación de analistas de laboratorio (factor operadores) y variación del punto muestreo de la red de suministro de un acueducto, sobre la variable respuesta del parámetro de alcalinidad total expresada en mg/L de CaCO_3 ; para ello se contó con las mediciones realizadas del ensayo de alcalinidad total por cuatro operadores en dos puntos de muestreo de la red de distribución (el punto más cercano a la planta de potabilización y el punto más lejano), durante los meses de enero a abril del año en curso. Posteriormente se realizó la validación de supuestos, obteniendo como resultado que los datos analizados obedecen a una distribución normal y cuentan con varianzas iguales. En cuanto a la evaluación de efectos sobre la variable respuesta realizado por medio del diseño experimental de estructura factorial (ANOVA factorial de efectos fijos 4X2), se obtuvo que el factor operadores y el factor punto de muestreo, no influyen de manera significativa (0,05) sobre la variable respuesta (alcalinidad total en mg/L de CaCO_3)

Palabras Clave:

Diseño experimentos, alcalinidad total, diseño factorial, efectos fijos, ANOVA

****Profesionales estudiantes especialización Estadística Aplicada**

FACTORIAL DESIGN 4X2 FIXED EFFECTS TO MEASUREMENT EVALUATION ON THE TOTAL ALKALINITY. CASE STUDY

SUMMARY

Multi-factor analysis is frequently used in the evaluation of various incidents factors on a same response variable, in this particular case is studied the possible of effects when exists variation of laboratory analyst (operators factor) and variation of sampling points in a supply red aqueduct, on the response variable for the parameter of total alkalinity in mg / L CaCO_3 ; we have measurements of total alkalinity test by four operators at two points sampling distribution red (near point and far point to the water treatment plant) in the months January to April this year. Subsequently we realized the assumptions validation, resulting, normal distribution and equal variances for the analyzed data. For the effects evaluation on the response variable, we development experimental design whit factorial structure (factorial ANOVA 4X2 Fixed Effects) and we obtained that factor operators and factor sampling point not significantly influence (0, 05) on the response variable (total alkalinity in mg / L of CaCO_3)

Keywords:

Experiments design, total alkalinity, factorial design, fixed effects, ANOVA

**** Professional Students Specialization Applied Statistics**

Capítulo 1. Introducción

El diseño de experimentos, se concibe como una herramienta estadística para inferir si una determinada variable respuesta, se ve afectada o no, por una o más fuentes de variabilidad, visto de otra forma, el diseño experimental se establece con base en un conjunto de pruebas experimentales de tal manera, que al analizar los datos, se pueda tomar decisiones respecto a un proceso (Box, Hunter & Hunter, 1978, citado por, De La Vara, 2002).

Dentro de los modelos de diseño de experimentos, se puede encontrar estructuras factoriales, las cuales enfocan la evaluación de los efectos de cada uno de los factores que influyen en una misma variable respuesta, y además permiten determinar la existencia de los efectos de interacción entre dos o más factores.

En el campo de la metrología, los resultados de las mediciones están sujetos a la influencia de fuentes de variación tales como: condiciones ambientales, instrumentos de medición, operadores, método de medida, patrones entre otros (Centro de Ingeniería de la Calidad, 2016). Por lo cual, la confiabilidad de los resultados puede verse en entre dicho, cuando no hay control sobre estos factores. Las técnicas de diseño estadístico de experimentos pueden utilizarse como una herramienta para evaluar dicha influencia de factores de variabilidad.

Uno de los factores que más llama la atención en el aseguramiento de la confiabilidad de las mediciones en metrología, es la presencia de errores humanos. Díaz, Sobrero y Pica (2004), afirman que no es suficiente utilizar métodos aprobados y estandarizados para lograr buenos resultados, los errores de los analistas pueden acarrear inexactitudes substanciales.

Un caso particular es la medición de parámetros de calidad en el agua potable, específicamente, la alcalinidad total, que se define de acuerdo a Severiche, Castillo y Acevedo (2013), como la capacidad para neutralizar ácidos, obedeciendo a la suma de todas las bases titulables (hidróxidos, carbonatos y bicarbonatos), la cual se expresa en mg/L de CaCO_3 .

La alcalinidad total, se determina generalmente a través de titulaciones volumétricas, lo que implica, que la medición puede llegar a depender de la competencia técnica de los analistas (operadores), ya que se debe apreciar el viraje del indicador empleado en el ensayo.

Adicionalmente, las determinaciones de alcalinidad habitualmente se utilizan en la interpretación y el control de procesos de tratamiento de aguas en general (González, 2015); no obstante, si se enmarca la alcalinidad total, como parámetro de calidad de agua potable; el decreto 1575 de 2006, artículo 9, estipula que la calidad de esta, se debe garantizar en cualquiera de los puntos que conforman el sistema de suministro, lo que conlleva a que los resultados de estos ensayos y demás, deban ser confiables.

En este caso se evidencian, entonces, dos potenciales factores trascendentales que pueden llegar a tener influencia sobre los resultados de las mediciones del parámetro de alcalinidad total: operadores y puntos de muestreo de suministro.

De esta forma, se propone utilizar un diseño con estructura factorial para evaluar la posible existencia de efectos causados por la variación de analistas que llevan a cabo los ensayos en mención y por la variación de puntos de muestreo de suministro del sistema de distribución, sobre la alcalinidad total (mg/L de CaCO_3).

Objetivo General

Evaluar los efectos de la variación de analistas y el punto de muestreo sobre la medición de la alcalinidad total, utilizando una estructura de diseño factorial.

Objetivos Específicos

Establecer el modelo de efectos para la alcalinidad total dependiendo de los factores influyentes.

Identificar si existe una combinación operario-punto de muestreo que pueda considerarse como factor de riesgo en la confiabilidad de la medición.

Capítulo 2. Marco de Referencia

El agua por ser vital en nuestro día a día, requiere de cierto aseguramiento de calidad para evitar riesgos que perjudiquen de alguna manera la salud. Para ello, es necesario realizar análisis de tipo fisicoquímicos que demuestre que el agua efectivamente puede ser de consumo humano y que no representa peligro alguno; entre éstos análisis encontramos los de alcalinidad total, turbiedad, dureza, conductividad, y demás (Páez, 2013).

Haciendo hincapié en el análisis que suscita este trabajo, es importante definir en primera instancia lo que a nivel global representa hablar de alcalinidad total, de acuerdo con Goyenola (2007, pág. 1), podemos interpretarla como: “la capacidad del agua para neutralizar ácidos y representa la suma de las bases que pueden ser tituladas, también desempeña un rol principal en la productividad de cuerpos en aguas naturales sirviendo como una fuente de reserva de CO_2 para la fotosíntesis”

De manera similar, (Vázquez y Manzano, 2009) refieren que es la cantidad de cambio en pH que se produce al adicionar ciertas cantidades de ácido, lo que garantiza que el agua sea suministrada o no, recordemos además que el agua natural posee iones de carbonato y bicarbonato, siendo grandes contribuyentes en este proceso.

Por otra parte, los límites razonables de la alcalinidad total están entre 20 mg/L y 250 mg/L expresada como CaCO_3 (el valor máximo para el territorio colombiano es de 200 mg/L), si se obtienen valores diferentes a éstos, los niveles de pH serían incorrectos y al contacto con el ser humano repercutiría en nuestro cuerpo evidenciando picazón en los ojos, irritación en la piel, entre otros, siendo éstas algunas razones que catalogan la alcalinidad junto con el calcio, hierro, zinc, y demás, características químicas que acarrear complicaciones tanto económicas como indirectas en la salud humana (Resolución 2115, 2007).

En la misma dirección de los resultados del presente estudio, se señala que las entidades, en este caso las plantas encargadas del tratamiento del

agua, deben contar con personal calificado y certificado bajo las competencias laborales las cuales están sujetas a las Resoluciones No.s 1076 de 2003 y 1570 de 2004, así como también se encuentran en la obligación de suministrar los equipos y materiales necesarios para realizar los ensayos requeridos (Resolución 2115, 2007).

Por otra parte, al realizarse el tratamiento de aguas y las mediciones de alcalinidad total correspondientes, el decreto 1575 de 2006, artículo 9, estipula que la calidad de ésta, se debe garantizar en cualquiera de los puntos que conforman el sistema de suministro, siendo esta una razón importante para seleccionar los puntos de muestreo más cercanos y lejanos al de partida, para así garantizar la confiabilidad de los datos recolectados.

Aquí puede evidenciarse dos puntos importantes que requieren de atención y es la influencia que hay entre los operadores y los puntos de muestreo sujetos a la evaluación de la alcalinidad total; desde el ámbito estadístico el hecho de hacer seguimiento a un conjunto de datos obtenidos a partir de réplicas, se le abona el nombre de diseño experimental, en dónde se busca optimizar la variable respuesta.

Capítulo 3. Marco Teórico

3.1 Diseño de experimentos

Cuando se realiza un experimento se hace énfasis en una particularidad y es la de descubrir algo acerca de un proceso, en ese sentido, puede definirse entonces como una prueba o serie de pruebas en las que se hacen modificaciones a las variables de entrada de un sistema para identificar las razones de los cambios que puedan darse en la respuesta de salida, de tal forma que el proceso sea afectado en forma mínima por fuentes de variabilidad externas (Montgomery, 2004).

El diseño de experimentos tiene una función particular y puede aplicarse de acuerdo a las necesidades tanto del experimento como del experimentador. A continuación, se citan algunas características que refieren a la importancia de este enfoque, los modelos que gestiona y como se utilizan en un contexto general.

3.2 Función de un diseño experimental

En el diseño de experimentos se efectúa una combinación de métodos, personas e incluso recursos que generan cambios a cierta entrada en una salida que tiene una o más respuestas observables, es así como se ha de observar que algunas variables del proceso son controlables, mientras que otras son no controlables, esta última depende de la finalidad de la prueba (Montgomery, 2004).

A continuación se presenta cinco objetivos que debe cumplir un experimento, Montgomery (2004, pág. 2):

- Determinar cuáles variables tienen mayor influencia sobre la respuesta y
 - Determinar cuál es el ajuste de las x que tiene mayor influencia para que y este casi siempre cerca del valor nominal deseado.
 - Determinar cuál es el ajuste de las x que tiene mayor influencia para que la variabilidad de y sea reducida.

- Determinar cuál es el ajuste de las x que tiene mayor influencia para que los efectos de las variables no controlables sean mínimos.
- Los experimentos se componen de varios factores, por lo que el experimentador deseara determinar la influencia que tienen esos factores sobre la respuesta de salida del sistema.

3.2.1. Modelo unifactorial o de enfoque de un factor a la vez

Consiste en seleccionar un punto de partida o una base de los niveles, para después variar cada factor en su rango dejando constantes los factores restantes en el nivel base y contrastando las gráficas que representan el cómo la variable respuesta fue afectada al variar cada factor dejando los demás constantes (Montgomery, 2004).

En este modelo se encuentran los experimentos que sólo tienen un factor de interés y a partir de éste hay varios niveles de estudio. El análisis de la varianza juega un papel importante aquí, ya que a partir de él se determina la diferencia entre las medias de los tratamientos siendo éstos variables aleatorias. Este modelo se reconoce fácilmente, ya que guarda la siguiente estructura:

Tabla 1. Datos típicos de un experimento de un solo factor

Tratamiento (Nivel)	Observaciones			Totales	Promedios
1	y_{11}	y_{12}	...	y_1	\bar{y}_1
2	y_{21}	y_{22}	...	y_2	\bar{y}_2
\vdots	\vdots	\vdots	...	\vdots	\vdots
a	y_{a1}	y_{a2}	...	y_a	\bar{y}_a

Fuente: Montgomery Douglas (2004), Diseño y Análisis de experimentos. Pág. 63

Del cuadro anterior se obtiene la formulación de un modelo unifactorial expresado de mediante el siguiente modelo lineal:

$$Y_{ij} = \mu_i + \tau_j + \varepsilon_{ij}$$

Ecuación 1. Modelo Unifactorial

Donde,

Y_{ij} : j- ésima observación del i-ésimo tratamiento

μ_i : Hace referencia a las medias totales

τ_j : Efecto del i- ésimo tratamiento

ε_{ij} : Error aleatorio

El modelo anterior puede verificarse mediante pruebas de hipótesis, dónde:

$$H_0: \mu_1 = \dots = \mu_a \text{ (el factor no influye)}$$

H_a : algún factor es diferente a un nivel de significación α (el factor influye)

En ese sentido, los errores se toman como variables aleatorias distribuidas normal e independientemente con media 0 y varianza σ^2 , siendo la última constante para todos los niveles del factor, para ser evaluada mediante un ANOVA como el que se muestra en la Tabla 2 (Montgomery, 2004).

Tabla 2. Tabla de análisis de varianza para el modelo con un solo factor y efectos fijos.

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F _o
Entre los tratamientos	$SS_{Tratamientos}$ $= n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2$	$a - 1$	$MS_{Tratamientos}$	$F_o = \frac{MS_{Tratamientos}}{MS_E}$
Error (dentro de los tratamientos)	SS_E $= SS_T - SS_{Tratamientos}$	$N - a$	MS_E	
Total	$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$	$N - 1$		

Fuente: Montgomery Douglas (2004), Diseño de experimentos. Pág. 70

3.2.2 Modelo factorial

Dentro del Diseño experimental se puede encontrar la estructura Factorial, esta consiste en realizar todas las combinaciones posibles de los niveles de varios factores, allí se describen dos tipos: cuantitativos y cualitativos, el primero hace referencia a valores métricos y los segundos a categorías de los factores,

Por otra parte, Montgomery (2004) explica que trabajar con varios factores es una estrategia experimental en la que los factores hacen variar en conjunto, en lugar de uno a la vez, este concepto es de vital importancia ya que permite ver el efecto de varias variables al tiempo sobre la variable respuesta, lo que le hace ver como el más eficaz de todos los diseños, puesto que se puede ver todas las interacciones de las posibles combinaciones en los niveles de los factores.

Este tipo de experimentos resultan eficaces ya que relaciona información sobre todos los factores, así como también se puede observar las respuestas de un factor dados diversos niveles de otro factor en un mismo experimento, así mismo ocurre interacciones entre los factores cuando su realización es independiente (Kuehl, 2001).

Se pueden encontrar tres tipos de efectos en un experimento factorial; éstos pueden ser: simples, principales y de interacción. Al referirse a los simples se hace referencia a las comparaciones entre los niveles de un factor a un solo nivel funcionando como contrastes; los principales son comparaciones entre los niveles de un factor donde se obtiene el promedio para todos los niveles de otro, por lo tanto se denotan como los efectos promedio de un factor y finalmente, los efectos de interacción que describen las diferencias entre los efectos simples (Kuehl, 2001).

3.2.3 Diseño factorial de dos factores

Es este el más simple de los diseños factoriales, está constituido por dos conjuntos de tratamientos, es decir, que hay " a " niveles del factor "A" y " b " niveles del factor "B", además las réplicas de los tratamientos contienen a su

vez las posibles combinaciones entre los niveles “ab”. Este es un diseño de filas por columnas, éste se presenta en la Tabla 3 (Montgomery, 2004).

Tabla 3. Arreglo general de un diseño factorial de dos factores

		FACTOR B			
		1	2	...	b
FACTOR A	1	$y_{111}, y_{112}, \dots, y_{11n}$	$y_{121}, y_{122}, \dots, y_{12n}$		$y_{1b1}, y_{1b2}, \dots, y_{1bn}$
	2	$y_{211}, y_{212}, \dots, y_{21n}$	$y_{221}, y_{222}, \dots, y_{22n}$		$y_{2b1}, y_{2b2}, \dots, y_{2bn}$
	...				
	a	$y_{a11}, y_{a12}, \dots, y_{a1n}$	$y_{a21}, y_{a22}, \dots, y_{a2n}$		$y_{ab1}, y_{ab2}, \dots, y_{abn}$

Fuente: Montgomery Douglas (2004), Diseño de experimentos. Pág. 176

Por otra parte, de la Tabla 3 se puede deducir el modelo de las medias con respecto a un diseño factorial:

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

Ecuación 2. Diseño Factorial - Modelo de las medias

Y de forma general, cuando el diseño consta de dos factores, el modelo matemático puede representarse mediante la ecuación:

$$y_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{ik} + \varepsilon_{ijk}$$

Donde:

y_{ijk} : Representa la variable respuesta por el sujeto i bajo el j -ésimo nivel del factor A y el k -ésimo nivel del factor B.

μ : Media general de la variable respuesta

α_j : Efecto principal asociado al tratamiento del j -ésimo nivel de la variable A.

β_k : Efecto principal asociado al tratamiento del k -ésimo nivel de la variable B.

$\alpha\beta_{ik}$: Efecto producido por la interacción entre el j -ésimo nivel de la variable A y el k -ésimo nivel de la variable B.

ε_{ijk} : Terminio de error aleatorio del modelo.

La variación total obtenida a partir del modelo general, consta de dos componentes, el de variación intragrupos (cuantifica la dispersión de los valores de la muestra con respecto a sus medias) e intergrupos (cuantifica la dispersión de las medias de las muestras con respecto a la media general) (Balluerka & Vergara, 2002).

En este caso se sugiere usar software estadístico para realizar los cálculos del ANOVA multifactorial (Tabla 4), de ese modo, se facilita probar la significancia de los dos efectos y sus respectivas interacciones, para ello se ha de visualizar los datos obtenidos en la columna F_o de la tabla de análisis de varianza, donde se puede decir que los datos no apoyan la hipótesis nula si los cocientes obtenidos son grandes (Montgomery, 2004).

Tabla 4. Tabla de Análisis de varianza multifactorial

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_o
Tratamientos A	SS_A	$a - 1$	$MS_A = \frac{SS_A}{a - 1}$	$F_o = \frac{MS_A}{MS_E}$
Tratamientos B	SS_B	$b - 1$	$MS_B = \frac{SS_B}{b - 1}$	$F_o = \frac{MS_B}{MS_E}$
Interacción	SS_{AB}	$(a - 1)(b - 1)$	$MS_{AB} = \frac{SS_{AB}}{(a - 1)(b - 1)}$	$F_o = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n - 1)$		
Total	SS_T	$abn - 1$		

Fuente: Montgomery Douglas (2004), Diseño de experimentos. Pág. 180

3.2.3.1 Pruebas de hipótesis sobre el efecto de los factores

Los efectos de los factores de carácter individual dependen de la ocurrencia o no de interacciones, para ello se debe definir la significancia de la interacción antes de determinar la de los efectos principales de los factores (Kuehl, 2001)

Hipótesis Nula para las interacciones:

$$H_o: (\alpha\beta)_{ij} = \bar{\mu}_{ij} - \bar{\mu}_i - \bar{\mu}_j + \bar{\mu} = 0 \text{ para toda } i, j$$

Hipótesis alternativa:

$$H_a: (\alpha\beta)_{ij} = \bar{\mu}_{ij} - \bar{\mu}_i - \bar{\mu}_j + \bar{\mu} \neq 0 \text{ para toda } i, j$$

3.3 Verificación de la adecuación del modelo

Para probar de manera formal que no hay diferencias entre las medias de los tratamientos, es necesario cumplir con unos supuestos, donde el modelo describe adecuadamente las observaciones y que además los errores siguen una distribución normal e independiente (Montgomery, 2004).

3.3.1 Supuesto de Normalidad

Para verificar este supuesto se puede realizar el histograma de los residuales, evidenciándose una distribución normal con centro en cero. Aun así, cuando se utilizan muestras pequeñas este método puede no ser muy útil, ya que no se logra evidenciar una violación seria de los supuestos.

3.3.2 Homogeneidad de Varianzas (Homocedasticidad)

Las pruebas estadísticas que aquí se usan tienen como objeto diagnosticar si existe desigualdad entre las varianzas de los residuales, para ello se consideran las siguientes hipótesis:

$$H_o: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$$

$$H_a: \text{El enunciado anterior no es verdadero para al menos un } \sigma_i^2$$

Una de las pruebas más usadas es la de Barlett, en la que el cálculo de su estadístico es aproximado a la distribución ji-cuadrada con $a-1$ grados de libertad cuando las a muestras son de poblaciones normales independientes, la H_o será rechazada cuando los valores de X_o^2 sean muy grandes. El valor P también puede ser usado para la toma de decisiones (Montgomery, 2004).

3.4 Comparaciones múltiples

En dado caso que las varianzas sugieran que las medias de los niveles de los factores son diferentes, se sugiere hacer comparaciones entre las medias individuales de los niveles para así encontrar diferencias puntuales, una prueba eficaz puede ser la de Tukey, ésta consiste en dejar un factor fijo a cierto nivel y ver cómo se comportan las interacciones dada dicha comparación (Montgomery, 2004).

Por otra parte, las comparaciones múltiples determinan los contrastes entre las medias que se requieren para profundizar en los efectos simples, es decir, que ayudan en la interpretación de los efectos significativos de la interacción (Kuelh, 2001).

Capítulo 4. Marco Metodológico

Los datos bajo estudio pertenecen al análisis fisicoquímico de muestras de agua potable, durante los meses de enero a abril del año en curso, específicamente de las mediciones realizadas para el parámetro de Alcalinidad Total por cuatro operadores, empleando el mismo instrumento e igual método. Las muestras tomadas pertenecen al punto de muestreo de distribución más cercano y más lejano, de una planta de potabilización de agua municipal del departamento de Cundinamarca.

4.1 Unidades experimentales:

Las unidades experimentales del presente trabajo son muestras de agua potable tomadas bajo las mismas condiciones de muestreo para garantizar la homogeneidad de las mismas. Para este caso específico las unidades experimentales son las mismas unidades observacionales.

4.2 Factores y niveles:

Se cuenta con dos factores denominados operador y punto de muestreo. El primer factor consta de cuatro niveles (operadores A, B, C y D) y el segundo factor está conformado por dos niveles (punto de muestreo cercano y lejano).

4.3 Variable Respuesta:

La variable respuesta refiere a la alcalinidad total expresada en unidades de mg/L de CaCO_3 . Correspondiente a los resultados de los análisis realizados a las unidades experimentales. La información relacionada a la variable respuesta, factores y niveles se presenta en la Tabla 5.

Tabla 5. Valores de Alcalinidad Total en mg/L de CaCO₃

Punto De Muestreo De Distribución		Operadores			
		A	B	C	D
	Cercano (PMC)	41	49	66	56
		55	54	63	39
		61	66	58	32
	Lejano (PML)	66	53	56	43
		54	58	55	63
		52	63	49	44

4.4 Tratamientos:

Los tratamientos representan las combinaciones de los niveles entre los operadores y los puntos de muestreo.

Como se precisa inferir si los operadores y los puntos de muestreo influyen de manera significativa sobre las mediciones de alcalinidad, se requiere un modelo de diseño de experimento para determinar dicha influencia sobre la variable respuesta, por lo tanto, se identifican algunos ítems experimentales (Tabla 6) a tener en cuenta para la ejecución adecuada de las pruebas correspondientes.

Tabla 6. Datos previos formación Diseño Factorial 4x2

ITEM	DESCRIPCIÓN
Tratamientos	8
Replicas:	3
Total de Mediciones	24
Modelo General Experimento	$Y_{IJ} = \mu + O + P + OP + \varepsilon_{ij}$
Hipótesis Nula Experimental	Los factores operador y punto de muestreo tienen los mismos efectos sobre la alcalinidad total.
Nivel de Significancia	$\alpha = 0,05$

4.5 Procedimiento

Para llevar a cabo el tratamiento estadístico sobre las unidades experimentales se hace uso del software libre R, versión 3.2.1 – 64 Bits (Paquete Rcmdr), donde inicialmente se ingresan los datos de acuerdo al esquema Anexo 1.

Una vez ingresados los datos al sistema, se realizan los pasos que se nombran a continuación:

- Diagramas exploratorios (Gráfico de medias y/o box-plot)
- ANOVA de un diseño con estructura factorial para identificar la significancia del operador, el punto de muestreo y su posible interacción.
- Validación de supuestos de normalidad y homocedasticidad bajo las pruebas de Shapiro Wilk y Bartlett respectivamente.
- Determinación del modelo
- Pruebas de comparación múltiple mediante la prueba de Tukey.

Capítulo 5. Análisis y Resultados

5.1 Diagrama exploratorio (Gráfico de medias)

Con el fin de dar una apreciación del comportamiento de los operadores en conjunto con el punto de muestreo, se emplea el siguiente gráfico (Figura 1).

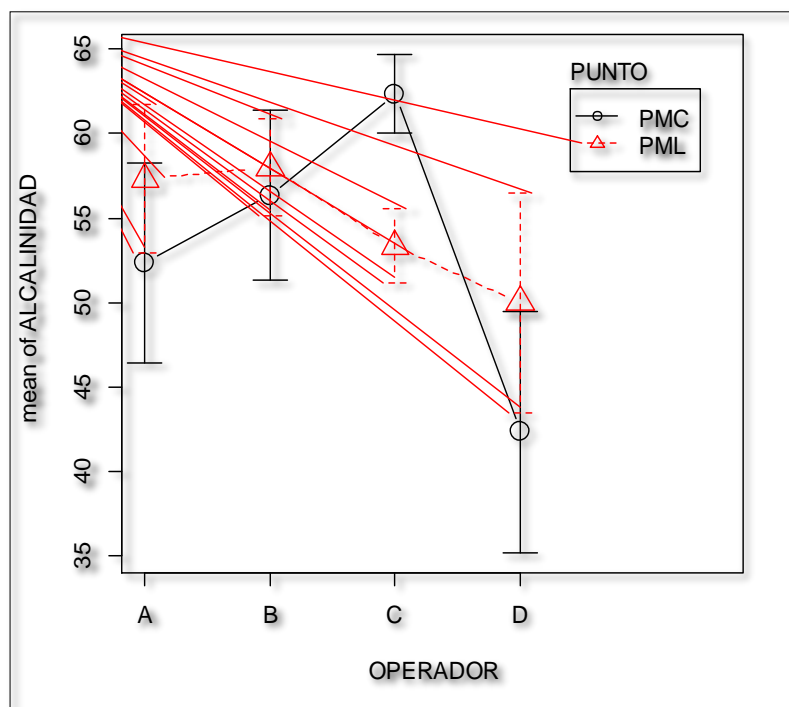


Figura 1. Gráfico de medias "Operador - Punto de Muestreo"

De la gráfica anterior se puede observar que los operadores C y D presentan medias característicamente diferentes, para la medición de alcalinidad en las muestras de agua potable del punto de muestro más cercano de la red de suministro. Adicionalmente el operador D presenta una mayor dispersión de las mediciones de alcalinidad para los dos puntos de muestreo.

5.2 ANOVA de un diseño con estructura factorial

El análisis de varianzas permite construir el modelo de acuerdo a los efectos que ejercen los factores frente a la variable respuesta, consecuentemente los resultados obtenidos se muestran en la Figura 2.

Anova Table (Type II tests)					
Response: ALCALINIDAD					
	Sum Sq	Df	F value	Pr(>F)	
OPERADOR	520.67	3	2.4189	0.1041	
PUNTO	10.67	1	0.1487	0.7049	
OPERADOR:PUNTO	240.67	3	1.1181	0.3711	
Residuals	1148.00	16			

Figura 2. Tabla de ANOVA estructura factorial

Por lo anterior, se puede inferir que el p- valor de cada uno de los factores es mayor al valor α , lo que demuestra que el factor operador, el factor punto de muestreo y las posibles interacciones de éstos, no influyen de manera significativa en la medición de alcalinidad total mg/L CaCO₃.

5. 3 Validación de supuestos

A continuación se muestra los resultados obtenidos de la validación del supuesto de normalidad y homocedasticidad, por medio del software R.

5.3.1 Normalidad

Para validar el supuesto de normalidad se recurre a la prueba de Shapiro – Wilk (Figura 3), la cual es muy útil cuando las muestras son pequeñas (Shapiro – Wilk, 1965), teniendo en cuenta las hipótesis nula y alternativa que se presentan a continuación:

H_0 = Los residuales se ajustan a una distribución normal

H_a = Los residuales no se ajustan a una distribución normal

```
Shapiro-Wilk normality test

data:  ALCALINIDAD
W = 0.94173, p-value = 0.1783
```

Figura 3. Resultado Prueba Normalidad Shapiro - Wilk, Software R

De acuerdo al resultado obtenido, se puede evidenciar que el p- valor es mayor al valor α , lo que implica el no rechazo de la Hipótesis nula (H_0), donde los residuales se ajustan a una distribución normal.

5.3.2 Homocedasticidad

La prueba de homocedasticidad se realizó mediante Bartlett y para corroborar dicho resultado se aplicó también Levene, en consecuencia se recurre a las hipótesis que se nombran a continuación:

H_0 = Las varianzas de los residuales intra tratamientos son iguales.

H_a = Las varianzas de los residuales no son iguales.

```
Bartlett test of homogeneity of variances

data:  ALCALINIDAD by interaction(OPERADOR, PUNTO)
Bartlett's K-squared = 4.378, df = 7, p-value = 0.7354

Levene's Test for Homogeneity of Variance (center = "median")
  Df F value Pr(>F)
group 7  0.3076 0.9401
    16
```

Figura 4. Resultado Prueba Homocedasticidad Bartlett, Levene.

Las pruebas de Bartlett y Levene (Figura 4), suscitan que el p- valor para cada una de ellas es mayor al valor α , por lo tanto, no se rechaza la H_0 que refiere a que las varianzas de los residuales son iguales.

5.4 Determinación del modelo

Una vez verificado el cumplimiento de los supuestos y el análisis de varianza se puede establecer que el modelo general (Tabla 6) se reduce a la siguiente expresión:

$$Y_{ij} = \mu + \varepsilon_{ij}$$

Ecuación 3. Modelo final

Donde,

μ : Representa la media general de los valores de alcalinidad total.

ε_{ij} : Es el error de la unidad experimental j al recibir el tratamiento i .

5.5 Prueba de comparaciones múltiples

El objetivo de esta prueba es determinar cuáles son los tratamientos con efectos diferentes (Figura 5); las pruebas de hipótesis dispuestas aquí, están en términos de las medias de los tratamientos, por ello cada una se define por un contraste.

Donde,

H_0 : Las medias de los tratamientos comparados son iguales.

H_a : Las medias de los tratamientos comparados no son iguales.

```
Fit: aov(formula = ALCALINIDAD ~ TRATAMIENTO, data = alcalinidad)
```

Linear Hypotheses:				
	Estimate	Std. Error	t value	Pr(> t)
APML - APMC == 0	5.0000	6.9162	0.723	0.995
BPMC - APMC == 0	4.0000	6.9162	0.578	0.999
BPML - APMC == 0	5.6667	6.9162	0.819	0.989
CPMC - APMC == 0	10.0000	6.9162	1.446	0.824
CPML - APMC == 0	1.0000	6.9162	0.145	1.000
DPMC - APMC == 0	-10.0000	6.9162	-1.446	0.824
DPML - APMC == 0	-2.3333	6.9162	-0.337	1.000
BPMC - APML == 0	-1.0000	6.9162	-0.145	1.000
BPML - APML == 0	0.6667	6.9162	0.096	1.000
CPMC - APML == 0	5.0000	6.9162	0.723	0.995
CPML - APML == 0	-4.0000	6.9162	-0.578	0.999
DPMC - APML == 0	-15.0000	6.9162	-2.169	0.417
DPML - APML == 0	-7.3333	6.9162	-1.060	0.956
BPML - BPMC == 0	1.6667	6.9162	0.241	1.000
CPMC - BPMC == 0	6.0000	6.9162	0.868	0.985
CPML - BPMC == 0	-3.0000	6.9162	-0.434	1.000
DPMC - BPMC == 0	-14.0000	6.9162	-2.024	0.497
DPML - BPMC == 0	-6.3333	6.9162	-0.916	0.980
CPMC - BPML == 0	4.3333	6.9162	0.627	0.998
CPML - BPML == 0	-4.6667	6.9162	-0.675	0.997
DPMC - BPML == 0	-15.6667	6.9162	-2.265	0.367
DPML - BPML == 0	-8.0000	6.9162	-1.157	0.933
CPML - CPMC == 0	-9.0000	6.9162	-1.301	0.886
DPMC - CPMC == 0	-20.0000	6.9162	-2.892	0.140
DPML - CPMC == 0	-12.3333	6.9162	-1.783	0.639
DPMC - CPML == 0	-11.0000	6.9162	-1.590	0.749
DPML - CPML == 0	-3.3333	6.9162	-0.482	1.000
DPML - DPMC == 0	7.6667	6.9162	1.109	0.946

Figura 5. Prueba Tukey - Comparaciones múltiples

De acuerdo a los resultados obtenidos en la Figura 5, el p_{valor} para cada uno de los contrastes es mayor al valor α , por ende no se rechaza la H_0 , lo que suscita que no existe diferencias significativas entre las medias de los dos tratamientos comparados.

En la figura 6 se muestran los diferentes intervalos de confianza obtenidos de la comparación de medias entre dos tratamientos.

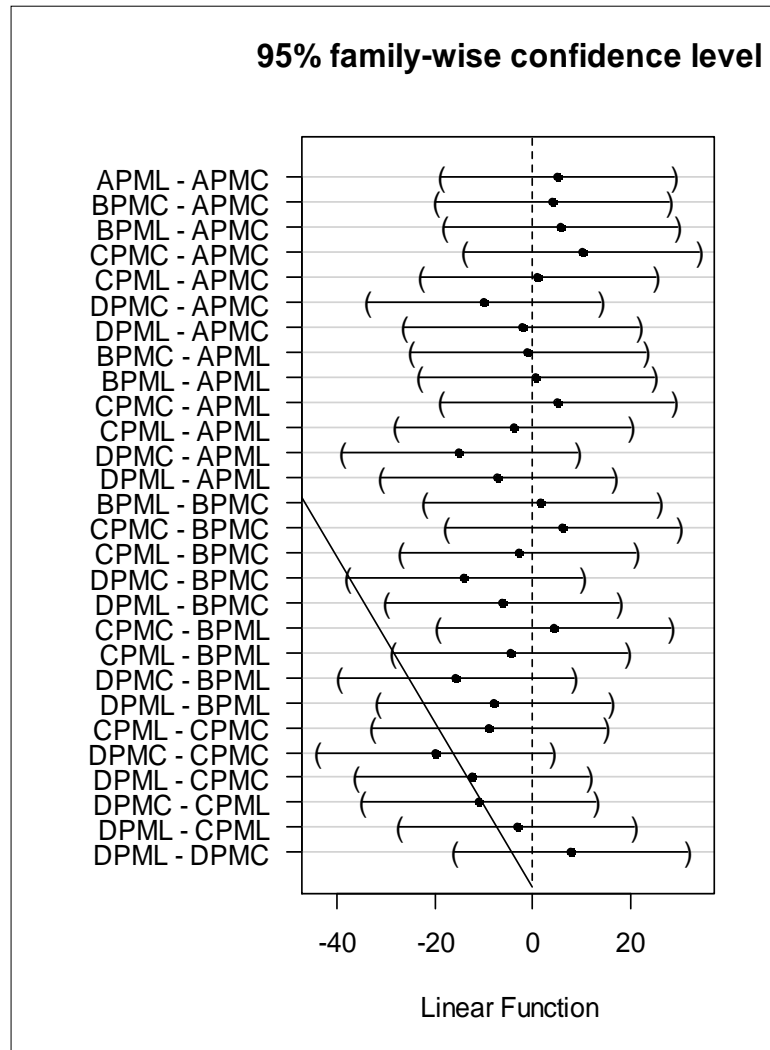


Figura 6. Intervalos de confianza de los tratamientos

Como el valor 0 pertenece a todos los intervalos, se puede afirmar que no existen diferencias significativas entre las medias de los tratamientos comparados.

A pesar de que todos los análisis expuestos anteriormente demuestran que no hay diferencias significativas entre los parámetros o efecto alguno de los factores estudiados al mismo tiempo sobre la variable respuesta, los diagramas exploratorios suscitan que de alguna manera puede existir influencia si se

evalúan los factores por separado, es decir, operador - variable respuesta y puntos de muestreo - variable respuesta.

5.6 Análisis Unifactorial

A partir de los diagramas de caja (Figura 7), puede observarse las interacciones individuales entre el operador – alcalinidad total y punto de muestreo – alcalinidad total.

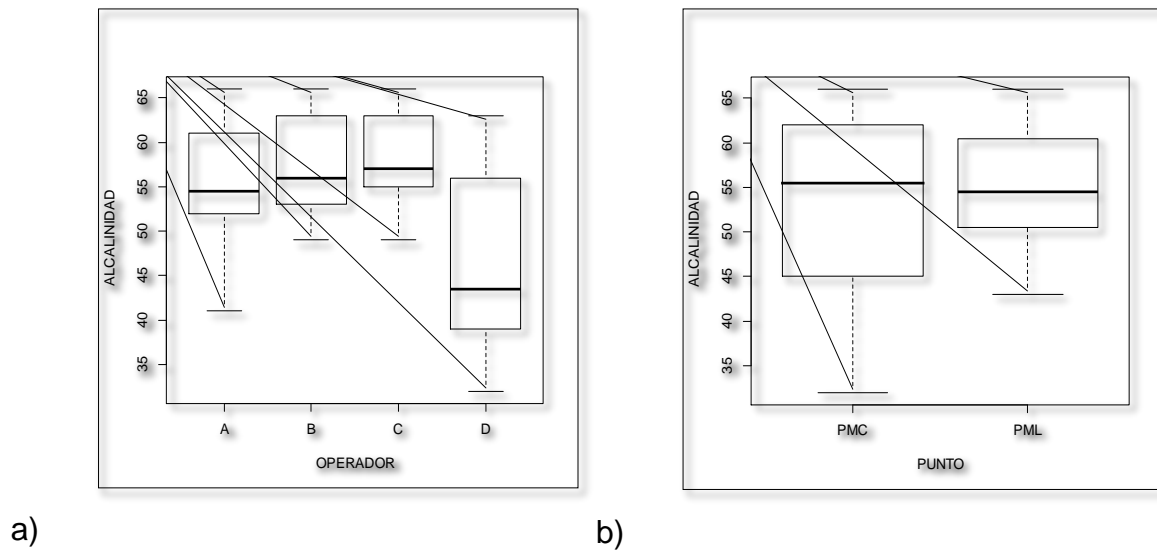


Figura 7. Diagramas de caja. a) Punto de Muestreo. b) Operador

Respecto a la gráfica anterior, se infiere que el operario D presenta mayor diferencia respecto a los otros, así mismo los puntos de muestreo son característicamente diferentes entre sí. Para verificar si estas diferencias son o no significativas, se realizan ANOVAS de un solo factor (Figura 8) para cada uno de los casos, donde las hipótesis a evaluar son:

H_0 : La medición de la alcalinidad total no se ve afectada por el operador

H_a : La medición de la alcalinidad total si se ve afectada por el operador

H_0 : La medición de la alcalinidad total no se ve afectada por el punto de muestreo

H_a : La medición de la alcalinidad total si se ve afectada por el punto de muestreo

```

> AnovaModel.8 <- aov(ALCALINIDAD ~ OPERADOR, data=alcalinidad)

> summary(AnovaModel.8)
          Df Sum Sq Mean Sq F value Pr(>F)
OPERADOR    3  520.7   173.56    2.481 0.0906 .
Residuals   20 1399.3    69.97
> AnovaModel.9 <- aov(ALCALINIDAD ~ PUNTO, data=alcalinidad)

> summary(AnovaModel.9)
          Df Sum Sq Mean Sq F value Pr(>F)
PUNTO       1   10.7    10.67    0.123  0.729
Residuals   22 1909.3    86.79

```

Figura 8. ANOVA de un solo factor para operador y punto de muestreo

De acuerdo con el resultado de la gráfica anterior, se evidencia que el p-valor sigue siendo mayor al valor α , por ende hay evidencia significativa para determinar que los factores no influyen sobre la variable respuesta.

De otro lado se evalúa el supuesto de homocedasticidad para cada uno de los factores (Tabla 8), donde se evidencia el cumplimiento del mismo, es decir, que las varianzas son iguales, ya que el p-valor es mayor al valor α .

Tabla 7. Resultados prueba Homocedasticidad de Factores Individuales

Prueba	Operador	P- Valor: Punto Muestreo
Bartlett	p-value = 0.4908	p-value = 0.1754
Levene	Pr(>F) 0.6623	Pr(>F) 0.2532

Adicionalmente, se realiza la prueba Tukey con el fin de determinar cuáles son los operadores con efectos diferentes (Figura 8); las pruebas de hipótesis dispuestas aquí, están en términos de las medias del factor operador, por ello cada una se define por un contraste.

Donde,

H_0 : Las medias del factor operador son iguales.

H_a : Las medias del factor operador son iguales.

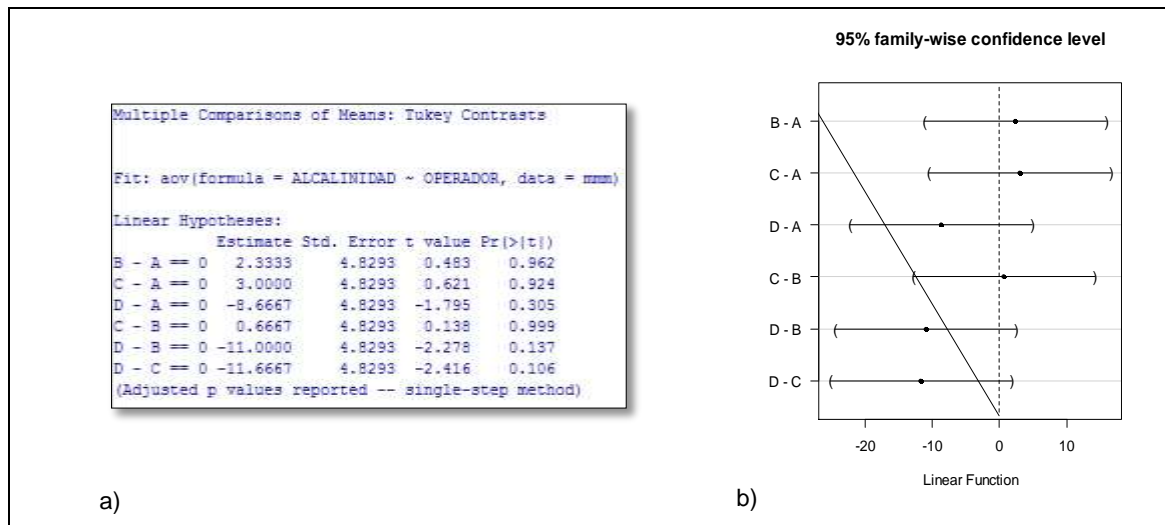


Figura 9. Prueba Tukey para el factor operador. a) Contrastes. b)

Intervalos de confianza.

De acuerdo a los resultados obtenidos en la Figura 9 para el caso a), el p-valor de cada uno de los contrastes es mayor al valor α , por ende no se rechaza la H_0 , lo que suscita que no existe diferencias significativas entre las medias del factor operador. Así mismo, en el caso b), se evidencia que el valor 0 pertenece a todos los intervalos, lo cual reafirma que no existen diferencias significativas entre las medias del factor operador.

Capítulo 6. Conclusiones

De acuerdo a los resultados obtenidos respecto al ANOVA FACTORIAL 4x2 se concluye que los dos factores evaluados (Operador y Punto de muestreo) tienen los mismos efectos sobre la variable respuesta (Alcalinidad total - mg/L CaCO₃).

Puesto que los factores (Operador y Punto de muestreo) tienen el mismo efecto sobre la variable respuesta, el modelo de efectos se ve reducido a la expresión:

$$Y_{IJ} = \mu + \varepsilon_{IJ}$$

Al tener el mismo efecto los factores en estudio (Operador y Punto de muestreo) sobre la variable respuesta, se infiere de igual forma que tampoco hay combinaciones de riesgo sobre la confiabilidad de la medición, es decir, independientemente del punto de muestreo del que provenga la muestra de agua potable y del operador que realice el análisis, se obtiene un valor de alcalinidad total similar.

Capítulo 7. Referencias

Balluerka, N. Vergara, A. (2002). Diseños de investigación experimental en Psicología: Modelos y Análisis de datos mediante el SPSS 10.0. Primer Edición. España. Pearson – Educación.

Centro de ingeniería de la calidad (2016). Estudio de repetibilidad y reproducibilidad R&R. Colombia. Recuperado De: <http://www.cicalidad.com/articulos/PAPER%20R&R.pdf>

De La Vara R., Diseños de Experimentos en Metrología (2002). México. Centro de Investigación en Matemáticas. Recuperado de: <https://www.cenam.mx/publicaciones/gratuitas/descarga/simposio%202002/doctos/te062.pdf>

Díaz M. C., Sobrero M. C y Pica Y. (2004). Aseguramiento y control de calidad de bioensayos. Ensayos toxicológicos y métodos de evaluación de calidad de aguas. Estandarización, intercalibración, resultados y aplicaciones. México. Instituto Mexicano de Tecnología del Agua/ Castillo Morales, Gabriela. Recuperado de: <https://books.google.com.co/books?id=GD7-N3154OIC&pg=PA125&lpg=PA125&dq=confiabilidad+de+los+resultados+en+laboratorios+de+ensayo&source=bl&ots=TognYNxov0&sig=einHHE2WPj5IPU t9WjbSthBzLM&hl=es-419&sa=X&ved=0ahUKEwim2Ky-tODMAhVKFx4KHY0YAwQ4ChDoAQgZMAA#v=onepage&q=confiabilidad%20de%20los%20resultados%20en%20laboratorios%20de%20ensayo&f=false>

González M. A. (2015). Estandarización De Las Metodologías Para El Análisis De pH, Alcalinidad Y Turbidez En Muestras De Agua Potable En El Laboratorio De La Asociación Municipal De Acueductos Comunitarios De Dosquebradas (AMAC). Colombia: Universidad Tecnológica De Pereira. Recuperado de: <http://repositorio.utp.edu.co/dspace/bitstream/11059/5107/1/6281G643.pdf>

Goyenola G., (2007). Guía para la utilización de valijas viajeras – Alcalinidad. Uruguay. Red de Monitoreo Ambiental Participativo de Sistemas Acuáticos. Recuperado de: http://imasd.fcien.edu.uy/difusion/educamb/propuestas/red/curso_2007/cartillas/tematicas/alcalinidad.pdf

Kuelh, R. (2001). Diseño de experimentos. Principios estadísticos de diseño y análisis de investigación. Segunda Edición. México D. F, México: Thomson – Learning.

Ministerio de la protección Social. Ministerio de Medio Ambiente, Vivienda y Desarrollo Territorial (2015). Resolución número 2115. Colombia. Recuperado de: <http://www.ins.gov.co/tramites-y-servicios/programas-de-calidad/Documents/resolucion%202115%20de%202007.MPS-MAVDT.pdf>

Montgomery, D. (2005). Diseño y Análisis de experimentos. Segunda Edición. México D. F, México. Limusa Wiley.

Vázquez M. H., Manzano B. (2009). Alcalinidad del Agua. España, Madrid. Recuperado de: <http://alcalinidaddelagua.blogspot.com.co/>

Severiche C. A., Castillo M. E., Acevedo R. L. (2013). Manual de Métodos Analíticos para la Determinación de Parámetros Fisicoquímicos Básicos en Aguas. Colombia: Editado por la Fundación Universitaria Andaluza Inca Garcilaso para eumed.net. Recuperado de: <http://www.eumed.net/libros-gratis/2013a/1326/1326.pdf>

Shapiro S., Wilk M., An analysis of variance for normality (complete samples) (1965). Gran Bretaña. Biometrika. Recuperado de: <http://webpace.ship.edu/pgmarr/Geo441/Readings/Shapiro%20and%20Wilk%201965%20-%20An%20Analysis%20of%20Variance%20Test%20for%20Normality.pdf>

Anexos

Anexo 1

Esquema de datos de los Tratamientos ingresados en “R”

OPERADOR	PUNTO DE MUESTREO	TRATAMIENTO	ALCALINIDAD
A	PMC	APMC	41
A	PMC	APMC	55
A	PMC	APMC	61
A	PML	APML	66
A	PML	APML	54
A	PML	APML	52
B	PMC	BPMC	49
B	PMC	BPMC	54
B	PML	BPMC	66
B	PML	BPML	53
B	PML	BPML	58
B	PML	BPML	63
C	PMC	CPMC	66
C	PMC	CPMC	63
C	PMC	CPMC	58
C	PMC	CPML	56
C	PML	CPML	55
C	PML	CPML	49
D	PMC	DPMC	56
D	PMC	DPMC	39
D	PMC	DPMC	32
D	PMC	DPML	43
D	PML	DPML	63
D	PML	DPML	44

Anexo 2.

Rutas empleadas en “R Commander”

Para la construcción del gráfico de medias de los factores se sigue la ruta:

Graficas =>Gráficas de medias =>Selección factores: Operador y Punto de Muestreo.

Para la construcción del gráfico de box plot de los factores se sigue la ruta:

Graficas =>Diagrama de Caja =>Gráfica por grupos => Seleccionar factor de estudio (Operador o Punto de Muestreo).

Para validar el supuesto de normalidad utilizando el software libre “R Commander” y partiendo de la previa base de datos ingresada, se emplea la siguiente ruta: Estadísticos => Resúmenes => Test de normalidad de Shapiro-Wilk.

Para validar el supuesto de homocedasticidad utilizando “R Commander”, se emplea la siguiente ruta: Estadísticos => Varianzas => Test de Bartlett

Para determinar los efectos de los factores sobre la variable respuesta en cuestión se tiene en cuenta la siguiente ruta: Estadísticos => medias=> ANOVA de múltiples factores

Para evaluar las combinaciones de medias se realiza la prueba de Tukey de acuerdo a la siguiente ruta: Estadísticos => medias=> ANOVA de un solo factor =>, teniendo en cuenta que se debe seleccionar la etiqueta de “TRATAMIENTO” y la opción de “Comparación dos a dos”